AD-A076 316 TECHNICAL OPERATIONS INC ALEXANDRIA VA COMBAT OPERAT--ETC F/G 15/4
SOME TYPES AND DISTRIBUTIONS OF INTELLIGENCE ERRORS.(U)
JAN 58 J HODGSON
UNCLASSIFIED CORG-M-67 NL

1 OF 1
AD
A076316

END
DATE
FILMED
11-79
DDC

CORG

tech/ops

COMBAT OPERATIONS RESEARCH GROUP

CONARC
ORG

(14) CORG-M-67
30 January 1958

LEVEL II

AD A076316

(11) 30 Jan 58

(12) 33

(9) CORG MEMORANDUM rept.,

(6) SOME TYPES AND DISTRIBUTIONS OF INTELLIGENCE ERRORS

by

(10) James Hodgson







Operated under
DEPARTMENT OF THE ARMY CONTRACT
between
HQ. CONTINENTAL ARMY COMMAND
and
TECHNICAL OPERATIONS, INC.

402 731

# DISCLAIMER NOTICE

**DEPARTMENT OF THE ARMY**
ARI FIELD UNIT, BENNING
U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
P.O. BOX 2086, FORT BENNING, GEORGIA 31905

PERI-IJ

8 August 1979

SUBJECT: Shipment of Documents

Defense Documentation Center
Cameron Station
Alexandria, VA 22314
ATTN: Selection & Cataloging

The Documents in these shipments are approved for public release. The distribution is unlimited.

FOR THE CHIEF:

Alexander Nicolini

ALEXANDER NICOLINI
Major, Infantry
R&D Coordinator

THIS IS A WORKING PAPER

representing the opinion of the author, and not necessarily that of CONARC or TOL. This memorandum is preliminary and, therefore, subject to revision and expansion.

CONTENTS

# SUMMARY

This memorandum exploits some data collected in the course of a cancelled project. It is principally exploratory.

The analysis presented in this memorandum illustrates some of the manifold errors in military intelligence. The data are drawn from the actions of V US Corps in Normandy, France, between 6 and 13 June and on 30 June 1944. Reliable statistical inferences cannot be drawn because of the paucity of material collected in the allotted time.

The investigation consisted of an historical reconstruction of the actual German situation in Normandy, collection of available data from V Corps G2 records concerning the enemy, and statistical analysis of the matched data. The analysis is based on a four-dimensional intelligence model that considers each intelligence target possesses size, type, location, and name, and emphasizing that name is useful only for the light it throws on the structure of the enemy force and the changes in that force.

The statistical analysis indicates that there is a significant difference between the frequency of error from one day to the next. Time available for data collection, however, prohibited the time-series analysis which this result indicates desirable. The implications of the data seem to reinforce common sense and military experience; that is, if a retreating enemy breaks contact, it is difficult to know what action he is taking. The data of this study indicate that the frequency with which elements of units are identified as entire units and the frequency with which units are thought to be other units rises sharply when contact is broken or relaxed. (See Table 12.)

The distribution of hits and misses for the different dimensions of intelligence can be presented in tabular form as in Table 12. This table represents only infantry units in a defensive attitude. A set of such tables based on adequate data might provide a useful origin for extrapolation into future operations by war gamers and designers of command-post exercises.

SOME TYPES AND DISTRIBUTIONS
OF INTELLIGENCE ERRORS

I. PROBLEM

The problem is to define and measure the distribution of some types of intelligence errors.

II. PURPOSE

The purpose of this study is to provide quantitative guidance to war gamers in introducing "realistically" some types of intelligence errors into war games and to intelligence analysts in estimating the reliability of collected battlefield intelligence.

III. DISCUSSION

Two sets of symbols appear in this memorandum. The first set of symbols is used in the definition of the problem and the data analysis (See Section IV.). The second set of symbols (See Section VII.) was developed by Raymond Burros for his analysis of the World War II data which will be published as a CORG Report.

A. Definitions

$B$ = inventory of Blue units = $B_1, B_2, B_3, \ldots . B_n$

$R$ = inventory of Red units = $R_1, R_2, R_3, \ldots . R_m$

$B_B$ = Blue feedback from B ($B_B \xleftarrow{B} B$); that is, $B_B$ is the Blue G3 picture of the Blue situation.

$B_R$ = Red feedback from B ($B_R \xleftarrow{R} B$); that is, $B_R$ is the Red G2 picture of the Blue situation.

$R_R$ = Red feedback from R ($R_R \xleftarrow{R} R$); that is, $R_R$ is the Red G3 picture of the Red situation.

$R_B$ = Blue feedback from R ($R_B \xleftarrow{B} R$); that is, $R_B$ is the Blue G2 picture of the Red situation.

$$R = \sum_{m=0}^{M} R_m \qquad B = \sum_{n=0}^{N} B_n$$

Each $B_N$ and $R_M$ is characterized by dimensions t, s, l, and w where s = size, t = type, l = location, and w = name or numerical designation:

$$B_{(s,t,l,w)} \text{ and } R_{(s,t,l,w)}$$

For the purpose of making combat decisions the Blue commander is interested in the size, type, and geographical distribution of the opposing troops. This means a definition of the intelligence inventory in terms of s and t is sufficient where the limiting values of dimension l are defined as being within an area L.

$$R_B = \sum_{m=0}^{M} R\left(s_m, t_m\right) \tag{1}$$

In treating World War II data, R designates German units and $R_B$ designates US intelligence.

B. Assumptions

1. Given a known order of battle (inventory of names of enemy units) and some knowledge of enemy organization practices, the type of a unit can be inferred from its name.

2. If w is considered as a dependent variable of s and t, then w can be used to estimate s and t.

Let $\bar{s}_t$ G2 estimate of the average strength of an enemy unit of type $t$.

$w_t$ number of units whose name implies that they are of type $t$.

$R_{B(t)}$ strength of that subset of $R_B$ which contains all units of type $t$.

Then

$$R_{B(t)} = \bar{s}_t w_t \qquad (2)$$

C. Postulate

Items in the inventories of R and $R_B$ may be considered matched if they are at identical locations on either $s$ or $w$. The statement

$$R_m = R_{B_m}$$

is true if (3)

$$l_R = l_{R_B} \text{ or } w_R = w_{R_B}$$

D. Burros' Definitions

In Raymond Burros' analysis of World War II data only those $R_m$ were treated which were located in an area L in dimension l consisting of s, the US Corps zone of interest which is an area bounded on its four sides by (1) the "front line", (3) a line parallel to the front line at a distance of 30 km to the German rear and (2) and (4) the produced lateral operational boundaries of the US corps. Part of Burros' analysis (German West Wall) uses $R_R$ as an estimate of R. The remainder (Normandy) of Burros' analysis and this memorandum is based on a comparison of $R_B$ with an approximation of R reconstructed by historical methods.

Burros' analysis is illustrated by the Venn diagram, Figure 1, in which

A  German units present but not identified

B  German units present and correctly and completely identified

C  German units not present and incorrectly but completely (spuriously) identified

D  German units incompletely identified whether present or not

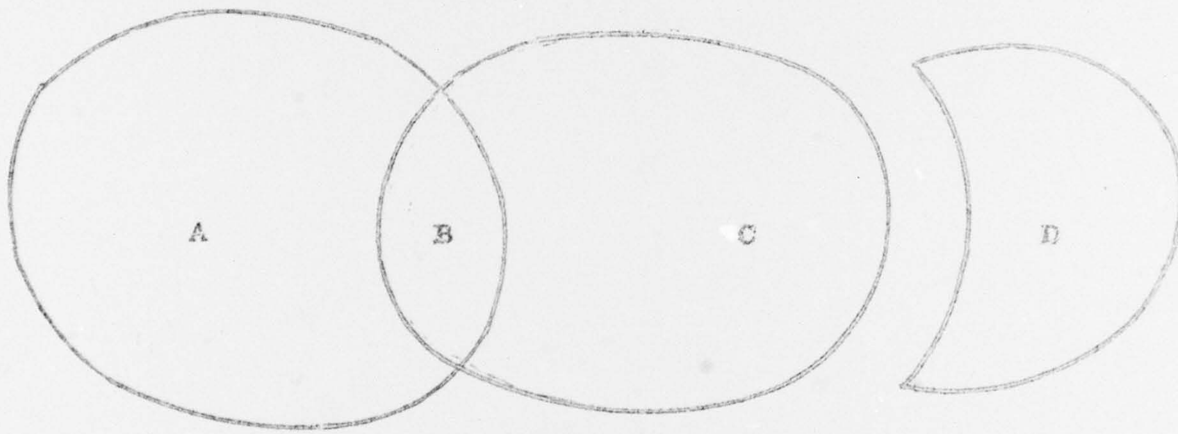


Figure 1. Venn diagram.

If German units are designated as Red units,

then

$$R = A \text{ and}$$

$$B_R = B + C + D$$

## IV. CATEGORIES OF INTELLIGENCE ERRORS

### A. Naming Errors, Errors in w.

Equations 2 and 3 indicate that - given the names of opposing units - estimates can be made of the size and type of the opposing force. The Burros' analysis contains deductions about the frequency with which such information was available to certain US commanders during some World War II actions. The following analysis looks at errors in s and t implied by errors in w.

#### 1. Error Types

Dimension w can be expanded into three dimensions. Components of w are number (name), series, and echelon.

Number (name) identifies which member of the indicated series is meant. The same number appears in different series. Exact identification requires knowledge of both. Where number is known but series is incorrect or unknown, matching must be done in dimension 1. (See par. c.)

Series implies t (type). Given number, the series may err in $n(n - 1)$ ways. (See par. a.)

Echelon implies s (size). Given a number, the echelon may also err in $n(n - 1)$ ways. (See pars. a. and b.)

##### a. Series Errors

The n series for the German army were army regimental infantry battalions, army separate infantry battalions, airforce field regimental infantry battalions, airforce field separate infantry battalions, SS regimental infantry battalions, SS separate infantry battalions, army engineer battalions, . . . .army regimental artillery battalions, . . . .army regimental tank battalions, etc. The number of the series exceeds 30, and $n(n - 1) = 30 (29) = 870$ possible series errors.

Since only the type is important, if these series are aggregated into four types in dimension t - infantry (including reconnaissance and engineers), artillery (including thin-skinned direct fire weapons and all indirect fire weapons), armor (including all thick-skinned vehicles mounting cannon), and non-combat units - then $n(n - 1) = 4\ (3) = 12$ possible errors in type.

Empirical Data

The only cases of such misidentification observed in Normandy in the V Corps G2 periodic reports were "large concentration of infantry, armor, and artillery" in the Forêt de Cerisy on 8 and 9 June (actually occupied by an artillery battalion (towed) and a divisional signal battalion) and some false reports of tanks. The data were scattered and too ill-defined to permit any reliable quantitative analysis.

b. Echelon Errors

<u>Pars Pro Toto</u> Confusions

The n echelons considered in the study are (1) company, (2) battalion, and (3) regiment. Then $n(n - 1) = 3\ (2) = 6$. Actually, the only two errors observed in the data were (a) companies present and thought to be battalions and (b) battalions present and thought to be regiments. These were actually identified units. The G2 made a "conservative," "cautious," or "pessimistic" estimate. Based on the data available, G2 could not be sure that the entire parent unit was not present. This type of error is designated a <u>pars pro toto</u> confusion.

Empirical Data

The number of such confusions is shown in Table 1.

Table 1.

| DAYS IN JUNE | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 30 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Company identified as Battalion | 1 | 0 | 0 | 0 | 13 | 11 | 8 | 3 | 0 | 36 |
| Battalion identified as Regiment | 0 | 1 | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 6 |
| Total | 1 | 1 | 0 | 0 | 16 | 13 | 8 | 3 | 0 | 42 |

The number of companies and battalions in V Corps zone of interest is given in Table 2.

Table 2.

| DAYS IN JUNE | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 30 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Companies | 36 | 27 | 51 | 48 | 60 | 63 | 96 | 102 | 60 | 543 |
| Battalions | 12 | 9 | 17 | 16 | 20 | 21 | 32 | 36 | 20 | 183 |

The percentage of cases which exhibited the two types of confusion on the different days is entered in Table 3.

Table 3.

| DAYS IN JUNE | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 30 |
|---|---|---|---|---|---|---|---|---|---|
| Company identified as Battalion | 2.78 | 0 | 0 | 0 | 21.67 | 17.67 | 8.33 | 2.94 | 0 |
| Battalion identified as Regiment | 0 | 0 | 0 | 0 | 15.00 | 9.52 | 0 | 0 | 0 |

By analysis of variance, it can be estimated at the 5-percent confidence level that there are differences between the frequency of confusion types which are not due to chance and further estimated at the 1-percent confidence level that the differences between the frequency of such confusions from one day to another is not due to chance alone. The results of the analysis are tabulated in Table 4.

Table 4.

| Source of variation | Sum of squares | Degrees of freedom | Mean square | F | $F_{.05}$ | $F_{.01}$ |
|---|---|---|---|---|---|---|
| Days | 743.2184 | 8 | 92.9023 | 14.61 | 3.44 | 6.03 |
| Type of confusion | 45.5230 | 1 | 45.5230 | 7.16 | 5.32 | 11.26 |
| Error | 50.8608 | 8 | 6.3576 | | | |
| Total | 839.6022 | 17 | | | | |

Since by nature a complete unit designation is given in each case of a <u>pars pro toto</u> confusion, and since these constitute a class of error, they belong in Burros' Category C.

c. Number Errors

(1) Type I

The errors in number are difficult to classify. There are those which fall into the class of typographical errors. Many permit interpretation to the extent that the original intent can be discerned with a high degree of reliability. There are others in which the erroneous numbers fit into the logical form of the numbering series but which are not represented by units actually in the area of attention. These are examined here.

Empirical Data, I.

Number errors made by V Corps G2 are summarized in Table 5.

Table 5.

| DAYS IN JUNE | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 20 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Separate Battalions | 0 | 0 | 1 | 0 | 4 | 4 | 2 | 0 | 0 | 11 |
| Regiments | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 |
| Total | 0 | 0 | 1 | 0 | 5 | 5 | 2 | 0 | 0 | 13 |

Since the data are meager, the analysis of variance has not been performed. It appears, however, by inspection that the results would tend to reinforce the confusion resulting from echelon errors.

(2) Type II

There is yet a third type of numbering error in which G2 recognizes the correct series, echelon, and location, but confuses the unit with a different unit in the corps zone of interest. These have been catalogued in terms of the resulting effect on the G2 picture of enemy locations and order of battle: The errors are incorrect spottings of units present.

V Battalions thought to be a different battalions of the same regiment (or another GHQ battalion in the same series)

W Battalions thought to be a different battalion of a different regiment in the same division

X Battalions thought to be a different battalion of a different division

Y Regiment thought to be a different regiment of the same division

Z Regiment thought to be a different regiment of a different division

$C_B = V + W + X$ Confusions resulting from misidentifications of battalions

$C_R = Y + Z$ Confusions resulting from misidentifications of regiments

$W + X + Y + Z$ Confusions resulting in misplacement of regiments

$X + Z$ Confusions resulting in misplacement of divisions

**Empirical Data, II**

The confusions are presented in tabular form in Table 6.

Table 6.

| June | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 30 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| V | 1 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 4 |
| W | 2 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 7 |
| X | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2 |
| Y | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Z | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2<br>15 |
| Sources | | | | | | | | | | |
| $C_B$ | 3 | 2 | 0 | 1 | 2 | 4 | 0 | 1 | 0 | 13 |
| $C_R$ | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2<br>15 |

The number of battalions and regiments present in the V Corps zone of interest is given in Table 7.

Table 7.

| June | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 30 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Battalions | 12 | 9 | 17 | 16 | 20 | 21 | 32 | 36 | 20 | 183 |
| Regiments | 4 | 3 | 5 | 5 | 10 | 9 | 8 | 10 | 8 | 62 |

To examine the risks inherent in pooling this data an analysis of variance is included in Table 8 which indicates no significant effects either of days or types of confusion.

Table 8.

| Source of variation | Sum of squares | Degrees of freedom | Mean square | F |
|---|---|---|---|---|
| Days | 39.2398 | 8 | 4.9050 | < 1.0 |
| Type of Confusion | 18.1491 | 4 | 4.5373 | < 1.0 |
| Error | 1302.0486 | 32 | 40.6889 | |
| Total | 1359.4327 | 44 | | |

With this assurance, all types of battalion misidentifications and all regimental misidentifications are pooled together in Table 9. The analysis of variance in Table 10 shows no significant difference between types of confusion or between days.

Table 9.

| June | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 30 |
|---|---|---|---|---|---|---|---|---|---|
| $C_B$/Bns | 25.0 | 22.0 | 0 | 6.3 | 10.0 | 19.0 | 0 | 2.8 | 0 |
| $C_R$/Regts | 0 | 0 | 0 | 0 | 10.0 | 11.1 | 0 | 0 | 0 |

Table 10.

| Source of variation | Sum of squares | Degrees of freedom | Mean square | F | $F_{.05}$ |
|---|---|---|---|---|---|
| Days | 607.1990 | 8 | 75.8999 | 1.58 | 3.44 |
| Type of confusion | 229.1376 | 1 | 229.1376 | 4.76 | 5.32 |
| Error | 385.3063 | 8 | 48.1632 | | |
| Total | 1221.6429 | 17 | | | |

2. Statistical Analysis, Difference Between Days

The data presented in this section indicate that errors in recognition of enemy units vary on different days. In the case of <u>pars pro toto</u> confusions, a significant difference was noted in the frequency between days. Analysis of the data in Table 4 indicates an intraclass correlation coefficient of 0.62 with 95-percent confidence limits 0.17 to 0.77, discussed in detail in Appendix A by Dr. Springer. This analysis indicates that a large part of the variance in the total frequency of such confusions is due to the difference between the frequency on different days.

The data represented in Tables 9 and 10 do not support a statistical hypothesis that such significant differences exist. However, the data as presented in Figure 2, while meager, do not refute the suspicion that errors in identification of German units occurred most frequently during periods when contact between V Corps and the Germans was being established or had been broken. Such periods are June 6-7 and June 10-11.

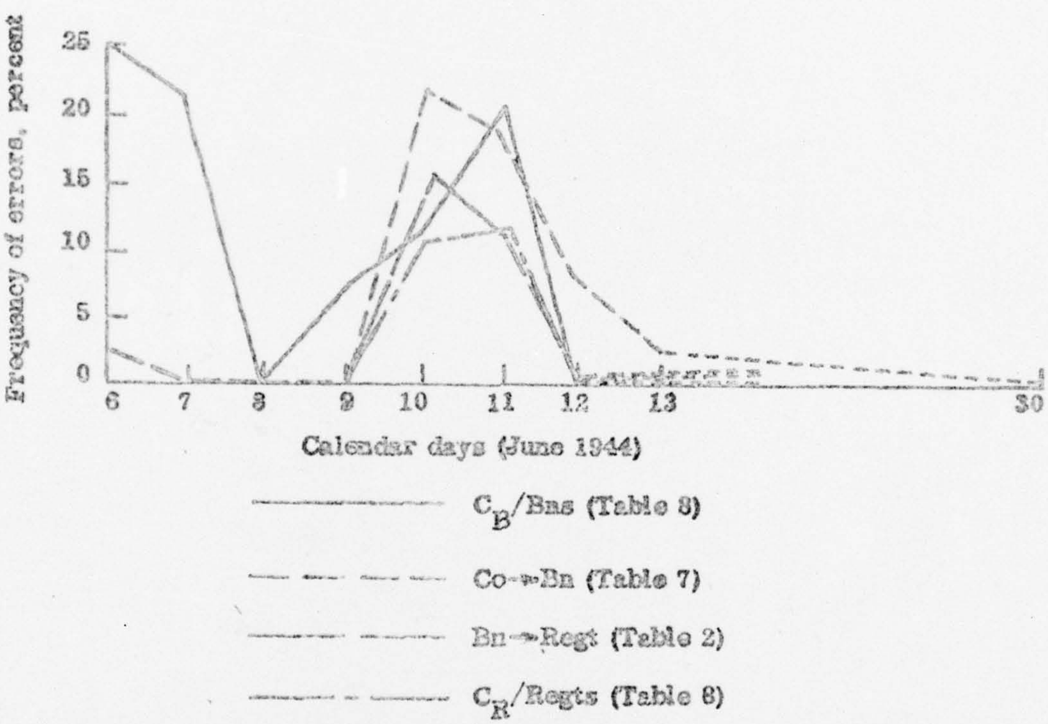


Figure 2. Frequency of errors plotted against calendar days.

The tendency toward smooth curves suggests that there might be interactions between days. The difficulty, of course, lies in defining the days in such a way that they can be grouped. This would require that the entire front of the corps be in the same state at the same time. The chances, however, of this happening are slight unless the corps is taking part in an over-all army push,

is being attacked all along its front by an opposing army, or the entire front is quiescent. These conditions are nearly satisfied in this instance. If portions of the front are active and other portions are inactive, there are an infinite number of populations of days which must be sampled.

In this particular case, the same state prevailed along the entire V Corps front on the night of 5/6 June (the day of the landing). The data are as of the close of the first day's fighting. With the withdrawal of 352d Infantry Division from defensive positions along the Aure River to the Elle during the night of 9/10 June, virtually the same state prevailed along the front; namely, V Corps troops were out of contact with the Germans. The resumption of the V Corps push southward took place on 11 June and created essentially the same situation that prevailed on 6 June.

If the data were to be subjected to a time-series analysis, the zero time would fall on 5 and 10 June.

If the entire intelligence problem were to be attacked on the basis of a time-series analysis, it would become quite costly. Every echelon and type of unit ($R_{(s, t)}$ for every state of combat (attack, defense, and quiescent) constitutes a separate series to the various degrees of identification. The analysis would have to be based on a best estimate of R and would be possible only where complete unit records are available for the intelligence targets. Such a project is economically feasible only where additional pay-offs can be obtained from the detailed reconstruction of R.[1]

---

[1]This should not be interpreted as implying that the time-series analysis is unfeasible. Such reconstructions of R in certain actions have been made and need exploitation through reconstruction of $R_{B}$. Reconstruction of R and B in real time is useful for a multitude of purposes.

B. Location Errors, Errors in l

Two types of location errors are possible. They are measurable in the first case between $R_{(w_1)}$ and $R_{B_{(w_1)}}$ and in the second case between $R_{B_{(s_1, t_1)}}$ and the nearest $R_{(s_1, t_1)}$, that is, the distance between the identified unit (whether the name is known or not) and the nearest actual unit of the same class. The second case is interesting only where unit identification is incomplete as to number. In a total of 342 command post days (all types and echelons) only 1 command post was located without knowledge of its name (0.29 percent per day). In a total of 183 battalion days, only 5 battalions (2.7 percent per day) and no regiments or higher units were located by type, size, and location without mention of name. The given point was on the main line of resistance and was occupied by Germans but not by an entire battalion. Since identification without knowledge of name is so rare, it seems advisable to restrict attention only to measurements of error of the first case. These are treated in Burros' study as radial miss distance.

Another type of location problem is related to the degree of precision of the location statement in dimension l. Since units occupy areas, their location can be indicated by a point, a line, or an area, representing increasing amounts of information and increasing precision of location. For the purpose of collecting data on V Corps intelligence in Normandy, use was made of the following categories.

1. Position Assertions

   a. Statements Independent of Line of Contact

      A Position of enemy unit given relative to other enemy units ("between," "behind"). There were no examples in the period selected.

B  Position of enemy unit given relative to US unit

Bag  opposite an army group $\{Ba \subset Bag\}$

Ba  opposite an army $\{Z \subset Ba \subset Bag\}$

Bc  opposite a corps {approximately equivalent to area Z}

Bd  opposite a division $\{Bd \subset Z \subset Ba\}$

Br  opposite a regiment $\{Br \subset Bd \subset Z\}$

Bb  opposite a battalion $\{Bb \subset Br \subset Bd \subset z\}$

b. Statements about Units in Contact

C  Point within area of unit ("vicinity," "behind").

D  Front line and boundaries specified.

c. Statements about Units Not in Contact

E  Location near a point ("vicinity")

F  Boundary of area occupied

2. Error Measurements

a. Percent of Units Identified

The identifications fraction represents the ratio between the sum over days of units identified and the sum over days of the units present. The percent identified is based on this fraction. This percent is further subdivided into the percent of total units present, identified according to each of the location statement categories.

b. Reliability

Each category of location statement was scored for reliability in terms of correctness or incorrectness with a measure of the magnitude of errors. The question whether Categories A, B, D, and F correctly represented the unit location can be answered "yes" or "no". The answers are evaluated respectively as 1.0 and 0. The reliability is the mean value of these answers.

Scoring for Categories C and E is somewhat more complicated, since they represent points within areas. In the case of front line units, V corps G2 normally designated a point not on the front line but close to it. The actual location was taken to be the nearest point on the front line. If the projection fell within the unit boundaries it was scored a hit; if outside, a miss. In the case of units not in contact, the location was scored a hit if the point identified lay within the unit boundaries.

c. Distance Errors

The measurement of errors was the distance from the actual but erroneous location statement to the nearest point on the map which would have received a correct score.

The significance of this error is as follows.

Let this distance be designated e

and

d diameter of the unit

r radius of effects

then by necessity

$$r = d + e$$

3. Empirical Data

TABLE 1 shows the V Corps intelligence product in terms of location statements. All data are pooled by type over days because there are insufficient data for more refined treatment. To attain a degree of generality, data from several corps would have been necessary. Conclusions can be only tentative and arrived at by inspection of the table.

a. V Corps G2 tended to use Category C for units in contact and Category E for units not in contact. These statements merely identify a point

within the area occupied by a unit. They contain little information.

b. Category D implies knowledge of unit boundaries. This was not used at all in the case of battalions and only in the case of 3 percent of regiments. This category implies more information than Category C. Knowledge of battalion boundaries implies more information than knowledge of regimental boundaries.

c. Not knowing the location of German division boundaries, V Corps G2 did not make location statements other than in Category B about divisions and corps.

## V. CONCLUSIONS

A. The model and the method of analysis both are workable, and might provide a means of treating objectively the injection of combat intelligence errors in war games and training exercises.

B. Though the data on which this memorandum was based were too meager to permit estimates of error frequencies by type, the data did reveal statistically significant differences between the frequency of errors on different days. This variation seems to vary systematically with the tactical situation as a function of the closeness of contact. Therefore, a modification of the model to determine error frequencies per day with days classified by prevailing tactical activity might represent a significant improvement.

Table 11. V US Corps G2 Intelligence product, OMAHA Beach, 6-13 June 1944.

| Intelligence target | Identification | | Location | | | Errors |
|---|---|---|---|---|---|---|
| | Fraction[a] | Percent | Category | Percent of total | Reliability | |
| Units in contact | | | | | | |
| Corps | 8/13 | 61 | Bc | 61 | 1.0 | |
| Divisions | 15/20 | 75 | Ba | 5 | 1.0 | |
| | | | Bc | 70 | 1.0 | |
| Regiments (Inf) | 26/38 | 68 | Ba | 3 | 1.0 | |
| | | | Bc | 5 | 1.0 | |
| | | | C | 58 | .82 | |
| | | | D | 3 | 0 | 5 $\pm$ 0, 0.3 km |
| Battalions (Inf)[b] | 37/109 | 34 | C | 34 | .5 | |
| (No armor committed during the period except separate companies.) | | | | | | |
| Units not in contact | | | | | | |
| Infantry Reserves | | | | | | |
| Divisions | 0/16 | 0 | | | | |
| Regiments | 0/19 | 0 | | | | |
| Battalions | 2/37 | 5 | C | 5 | .5 | 7 $\pm$ 2.07 km |
| Armor Reserves | | | | | | |
| Battalions | 0/6 | 0 | | | | |
| Artillery Battalions | | | | | | |
| Medium field | 2/8 | 25 | E | 13 | 1.0 | |
| | | | F | 13 | 1.0 | |
| Light field | 2/23 | 9 | E | 4 | 1.0 | |
| Antiaircraft | 9/56 | 16 | Bc | 16 | 1.0 | |

Footnotes to table on preceding page

[a] Identification fraction = $\frac{\text{Sum of units x days identified}}{\text{Sum of units x days present}}$

[b] The linear errors fall into three classes:

1. 40 km. An error in placing a battalion because of misinterpretation of the numbering system. The 9th Co., belonging to the 3rd Battalion of a Regiment was identified as belonging to the 2d Battalion. The company was correctly identified. The Order of Battle team made the error.

2. Errors arising from loss of contact. The 352d Division withdrew on the evening of 9/10 June to positions on the Elle River. V Corps did not maintain contact. V Corps identified stragglers as representing the division and was deceived by the active resistance of 17th SS Reconnaissance Battalion into thinking it was opposed in force.
   Error: (mean deviation from the mean) 18.4 ± 2.7 (8 cases)
   (median) 18
   18 km represents approximately the division front.

3. Normal errors: (mean deviation from the mean)
   2.7 ± 1.2 (12 cases)
   (median) 2.5

It is worthy of note that the mean deviation from the mean in B is about twice that of C. This seems reasonable in view of the fact that the division withdrew in formation and merely rearranged the battalions somewhat within the regiments but did not rearrange the order of the regiments in the front.

Table 11. V US Corps G2 Intelligence product, OMAHA Beach, 6-13 June 1944 (Concluded).

| Intelligence | Identification | | Location | | | Errors |
|---|---|---|---|---|---|---|
| | Fraction[a] | Percent | Category | Percent of total | Reliability | |
| Command Posts | | | | | | |
| Corps[b] | 2/17 | 12 | E | 12 | 0 | $27 \pm 0$ km |
| Divisions[c] | 5/20 | 25 | E | 25 | 0 | |
| Infantry Regiments (committed)[d] | 2/35 | 6 | E | 6 | .5 | |
| Infantry Regiments (reserve) | 0/19 | 0 | | | | |
| Infantry Battalions (committed) | 1/98 | 1 | E | 1 | 0 | $3.5 \pm 0$ km |
| Infantry Battalions (reserve) | 0/37 | 0 | | | | |
| Armor Regiments (reserve) | 0/4 | 0 | | | | |
| Armor Battalions (reserve) | 0/6 | 0 | | | | |
| Artillery Brigades[e] | 0/10 | 0 | | | | |
| Artillery Regiments[f] | 2/21 | 10 | E | 10 | 0 | $3.6 \pm 1.4$ km |
| Artillery Battalions[g] | 4/80 | 5 | E | 5 | .5 | |

a Identification fraction = $\frac{\text{Sum of units x days identified}}{\text{Sum of units x days present}}$

b On two successive days the command post of LXXXIV Corps (German) was carried as identified. The error was 27 km westward. The Corps command post was not subsequently identified.

c The errors were 18, 3, 3, 0.5, and 0.5 km. $6.2 \pm 3.9$ km. The largest error was associated with the loss of contact on 10 June. The other errors were normal: $3.3 \pm 2.9$ km.

d The erroneous location missed by 1 km, the other was within 100 m.

e These Artillery Brigade command posts correspond to US Corps Artillery headquarters.

f These Artillery Regiment command posts correspond to US Division Artillery headquarters.

g The two errors were both 30 km and occurred on 8 and 9 June.

TABLE 12. DISTRIBUTION OF ERRORS; $\left\{R_{B_m} \xleftarrow{B} R_m\right\}$ INFANTRY UNITS; $\{t\}$

| Intelligence target, $[s]^a$ | Location category, $[l]^a$ | Naming error, $[w]^a$ | Frequency, percent | Random number |
|---|---|---|---|---|
| Corps (6) | Opposite Army Group $X_B$ | none | 0 | 0 |
| | Opposite Army $X_B$ | none | 0 | 0 |
| | Opposite Corps $X_B$ | none | 61 | 1-61 |
| | Opposite Division $X_B$ | none | 0 | 0 |
| | None $\left[R_{B_m} = 0\right]$ | none | 39 | 62-100 |
| Divisions (13) | Opposite Army Group $X_B$ | none | 0 | 0 |
| | Opposite Army $X_B$ | none | 5 | 1-5 |
| | Opposite Corps $X_B$ | none | 70 | 6-75 |
| | Opposite Divisions $X_B$ | none | 0 | 0 |
| | None $\left[R_{B_m} = 0\right]$ | none | 24 | 76-100 |
| Regiments (62) | Opposite Army Group $X_B$ | none | 0 | 0 |
| | Opposite Army $X_B$ | none | 3 | 1-3 |
| | Opposite Corps $X_B$ | none | 3 | 4-6 |
| | Opposite Divisions $X_B$ | none | 7 | 7-13 |

TABLE 12. DISTRIBUTION OF ERRORS; $\left\{R_{B_m} \xleftarrow{B} R_m\right\}$ INFANTRY UNITS; {t} (Continued)

| Intelligence target, $[s]^a$ | Location category, $[l]^a$ | Naming error, $[w]^a$ | Frequency, percent | Random number |
|---|---|---|---|---|
| Regiments (62) (Continued) | Near $P_{(x, y)}$ correct | none | 30 | 14-43 |
| | Near $P_{(x, y)}$ incorrect | none | 15 | 44-58 |
| | Near $P_{(x, y)}$ | another regiment | 3 | 59-61 |
| | Boundaries given | none | 3 | 62-64 |
| | None $\left[R_{B_m} = 0\right]$ | none | 36 | 65-100 |
| Battalions (183) regimental (114) | Opposite Army Group $X_B$ | none | 0 | 0 |
| | Opposite Army $X_B$ | none | 0 | 0 |
| | Opposite Corps $X_B$ | none | 0 | 0 |
| | Opposite Division $X_B$ | none | 0 | 0 |
| | Near $P_{(x, y)}$ correct | none | 7 | 1-7 |
| | Near $P_{(x, y)}$ incorrect | none | 7 | 8-14 |
| | Near $P_{(x, y)}$ | identified as parent regiment | 4 | 15-18 |
| | Near $P_{(x, y)}$ | identified as undefined battalion in parent regiment | 2 | 19-20 |
| | Near $P_{(x, y)}$ | identified as other battalion in parent regiment | 0 | 0 |

TABLE 12. DISTRIBUTION OF ERRORS; $\left\{R_{B_m} \xleftarrow{B} R_m\right\}$ INFANTRY UNITS; {t} (Concluded)

| Intelligence target, $[s]^a$ | Location category, $[l]^a$ | Naming error, $[w]^a$ | Frequency, percent | Random number |
|---|---|---|---|---|
| **Battalions (182) regimental (114)** | Near $P_{(x, y)}$ | identified as battalion in other regiment, same division | 1 | 21 |
| (Continued) | Near $P_{(x, y)}$ | identified as battalion in other division | 2 | 22-23 |
| | Boundaries given | none | 77 a | 24-100 |
| | None $\left[R_{B_m} = 0\right]$ | none | 77 | 24-100 |
| **Divisional (46)** | Near $P_{(x, y)}$ correct | none | 11 | 1-11 |
| | Near $P_{(x, y)}$ incorrect | none | 9 | 12-20 |
| | Near $P_{(x, y)}$ | another battalion | 0 | 0 |
| | Boundaries given | none | 0 | 0 |
| | None $\left[R_{B_m} = 0\right]$ | none | 80 | 21-100 |
| **Separate** | Near $P_{(x, y)}$ correct | none | 15 | 1-15 |
| | Near $P_{(x, y)}$ incorrect | none | 15 | 16-30 |
| | Near $P_{(x, y)}$ | another battalion | 2 | 31-32 |
| | Boundaries given | none | 0 | 0 |
| | None $\left[R_{B_m} = 0\right]$ | none | 68 | 33-100 |

Footnotes to table on preceding pages

[a] The symbols given in square brackets refer to the terminology adopted and explained in the discussion which follows. The numbers in parentheses under s are the number of cases on which the frequency percentage is based.

The distribution of location errors is not treated in this memorandum since the data is covered on a broader basis in Raymond Burros' CORG Report. The location errors ranged from 0 to about 25 km. The distinction between correct and incorrect location is discussed on p. 15.

## APPENDIX A

## THE INTRACLASS CORRELATION COEFFICIENT

by

Dr. Melvin A. Springer

R. A. Fisher has introduced the notion of intraclass correlation as a measure of the variation of the mean values from class to class when $N$ variates (items) are grouped into $k$ classes of $n$ variates each. In order to define the intraclass correlation coefficient, we must regard the model for the analysis of this one-way classification as a Linear Hypothesis Model rather than a Components of Variance Model. If a Components of Variance Model is used, it is necessary to assume that for any class the variate is normally distributed with some mean and with a variance $\sigma^2$ which does not change from class to class. In order to employ the Linear Hypothesis Model, we must make a further assumption; this assumption is also necessary in order to define the intraclass correlation coefficient. In the Components of Variance Model, the mean values for different classes have been considered as fixed, unknown constants. We now consider an infinite population of different classes such that with each class is associated a normal distribution with a mean value $\mu$ and standard deviation $\sigma$ expressing the distribution of the observations within that class. Again, it is assumed that $\sigma$ is the same for all classes, but $\mu$ itself is a random variable normally distributed with a certain mean $\nu$ and standard deviation $\sigma^1$. The $k$ classes observed constitute a random sample of size $k$ from the population of classes, and the corresponding mean values $\mu_1, \mu_2, \ldots, \mu_k$ are a random sample from a normal distribution with a mean value $\nu$ and standard deviation $\sigma^1$. In this formulation, the null hypothesis $\mu_1 = \mu_2 = \ldots = \mu_k$ can be expressed as $\sigma^1 = 0$, since $\sigma^1 = 0$ means that the variable $\mu$ is really a constant.

If this one-way classification is analyzed by a Linear Hypothesis Model, the intraclass correlation coefficient in the population from which the k classes were drawn may be defined as

$$\rho_I = \frac{\sigma_1^2}{\sigma^2 + \sigma_1^2} \qquad (1)$$

so that testing the hypothesis $H_0$:

$$\mu_1 = \mu_2 = \ldots = \mu_k$$

is equivalent to testing the hypothesis $H_0^1$:

$$\rho_I = 0$$

A significant F implies that $\rho_I$ is significantly different from zero.

An unbiased sample estimate of the population intraclass correlation coefficient $\rho_I$ is

$$r_I = \frac{\dfrac{S_{\bar{x}}^2 - S^2}{n}}{S^2 + \dfrac{S_{\bar{x}}^2 - S^2}{n}} \qquad (2)$$

where

$$S_{\bar{x}}^2 = \frac{n \sum_{i}^{k} \left(\bar{x}_i - \bar{x}\right)^2}{k - 1}$$

$$S^2 = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} \left(\bar{x}_{ij} - \bar{x}\right)^2}{k(n - 1)}$$

Reducing Fraction 2 to its lowest terms and dividing the numerator and denominator of the reduced fraction by $S^2$

$$r_I = \frac{S_{\bar{x}}^2 - S^2}{S_{\bar{x}}^2 + (n-1)\, S^2} = \frac{\dfrac{S_{\bar{x}}^2}{S^2} - 1}{\dfrac{S_{\bar{x}}^2}{S^2} + (n-1)} \tag{3}$$

Since $\dfrac{S_{\bar{x}}^2}{S^2} = F$ (i.e., has the $F$ distribution), Equation 3 becomes

$$r_I = \frac{F - 1}{F + n - 1} \tag{4}$$

We are usually interested either in testing the hypothesis $\rho = \rho_0$ or in establishing a 100(1-p)percent confidence interval for $\rho_I$.

Consider the ratio $\lambda^2 = \dfrac{\sigma^{1^2}}{\sigma^2}$. $\rho_I$ may be written in the form

$$\rho_I = \frac{\lambda^2}{1 + \lambda^2} \tag{5}$$

The variable

$$F = \left(\frac{\dfrac{S_{\bar{x}}^2}{\sigma^2 + n\sigma^{1^2}}}{\dfrac{S^2}{\sigma^2}}\right) = \frac{S_{\bar{x}}^2}{S^2} \times \frac{\sigma^2}{\sigma^2 + n\sigma^{1^2}} \tag{6}$$

has the $F$ distribution with $k - 1$ and $k(n - 1)$ degrees of freedom. Dividing numerator and denominator of Equation 6 by $\sigma^2$, we have

$$F = \frac{S_{\bar{x}}^2}{S^2} \times \frac{1}{1 + n\lambda^2} = F^*\left(\frac{1}{1 + n\lambda^2}\right) \tag{7}$$

That is, we are now concerned with a new $F$ variable which consists of the original $F$ variable (denoted by $F^*$) multiplied by the factor $\dfrac{1}{1 + n\lambda^2}$. To test the hypothesis that $\rho_I = 0.60$ for days, we note from Snedecor's $F$ table that the 5-percent critical value for 8 and 9 degrees of freedom is $F_2 = 3.23$, while for 9 and 8 degrees of freedom it is $F_1^{\,1} = 3.40$. The two critical values we are interested in are

$F_1 = \frac{1}{F_1} = 0.29$ and $F_2 = 3.23$. We reject the hypothesis $\rho_I = 0.60$ if and only if the inequality

$$F_1 \le \frac{F^*}{1 + n\lambda^2} \le F_2 \tag{8}$$

is not satisfied; that is, if and only if $\frac{F^*}{1 + n\lambda^2}$ does not satisfy the condition

$$0.29 \le \frac{F^*}{1 + n\lambda^2} \le 3.23 \tag{9}$$

Since $\frac{F^*}{1 + n\lambda^2} = \frac{9.05}{1 + 9(1.5)} = 0.62$, the inequality (Equation 9) is satisfied and we do not reject the hypothesis $\rho_I = 0.60$

A 100(1-p)percent confidence interval for $\rho_I$ is

$$\left(\frac{\lambda_1^2}{1 + \lambda_1^2}, \frac{\lambda_2^2}{1 + \lambda_2^2}\right) \tag{10}$$

where

$$\lambda_1^2 = \frac{1}{n}\left(\frac{F^*}{F_2} - 1\right)$$

$$\lambda_2^2 = \frac{1}{n}\left(\frac{F^*}{F_1} - 1\right)$$

and where $F_1$ and $F_2$ are the lower and upper 100p-percent critical values of $F$. Thus, in the case of the intraclass correlation coefficient for days, we have

$$\lambda_1^2 = \frac{1}{9}\left(\frac{9.05}{3.23} - 1\right) = 0.20$$

$$\lambda_2^2 = \frac{1}{9}\left(\frac{9.05}{0.29} - 1\right) = 3.36$$

so that $\frac{\lambda_1^2}{1 + \lambda_1^2} = 0.17$ and $\frac{\lambda_2^2}{1 + \lambda_2^2} = 0.77$. Therefore, the 95-percent confidence interval for $\rho_I$ is (0.17, 0.77).

It should be pointed out that a significant $F$ in the above example indicates both that $\sigma_{Days}^2 \neq 0$ and that the associated intraclass correlation coefficient is

significantly different from zero. If that is all we are interested in, we may stop at this point. If, however, we desire a 95-percent confidence interval for $\rho_I$, we must proceed as indicated above. It is important to note that we cannot obtain said confidence interval by merely substituting $F_1$ and $F_2$, respectively, in the formula

$$r = \frac{F - 1}{F + n - 1}$$

Such a substitution in the days classification gives the interval (-0.09, 0.20), which clearly differs from the correct interval (0.17, 0.77).

In the case of the days classification, the *significant* $r_I$ means that $\sigma_{Days}$ (the variance between day means) constitutes a large part of the total variation $\sigma_{Days}{}^2 + \sigma_{Confusion}{}^2 + \sigma_{Error}{}^2$.

It should be mentioned that the lower endpoint $\frac{\lambda_1{}^2}{1 + \lambda_1{}^2}$ of the confidence interval, Equation 10, will be negative if $\frac{F^*}{F_2} < 1$. This would seem to indicate that $\sigma_1{}^2$ in Equation 1 may be negative, which is clearly impossible. Such a result is due to the fact that some control has made the sampling non-random. Either the observations within the columns have been spread artificially or the column means have been compressed into a range smaller than that ordinarily found in sampling.